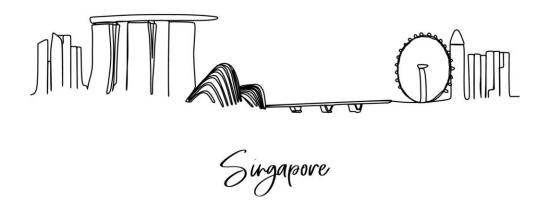
DR. ALVIN'S PUBLICATIONS

# LINEAR REGRESSION USING EXCEL

# BY DR. ALVIN ANG



# **CONTENTS**

Ι.	What is Linear Regression3
<i>II</i> .	Step 1: Scatter Plot to Check Linearity4
Α.	Alternative Way for Curve Fitting7
<i>III</i> .	Step 2: Regression Analysis10
IV.	Step 3: Analyzing the Regression Analysis Output12
Α.	Multiple R12
В.	R square14
C.	SSR / SSE / SST15
D.	Degrees of Freedom (df)17
E.	Adjusted R <sup>2</sup>
F.	Standard Error20
G.	Coefficients
H.	F Stat23
١.	Significance F
J.	t Stat27
к.	P-Value
L.	Lower and Upper 95%32
V.	Assumptions of Linear Regression33
VI.	Linear Regression By Hand34
Abo	ıt Dr. Alvin Ang

# **2** | P A G E

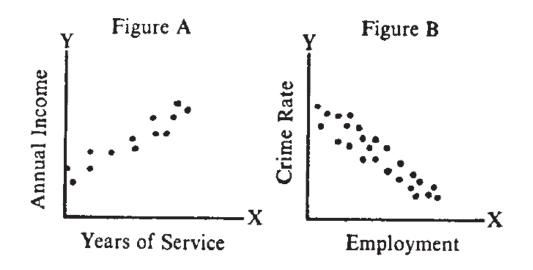
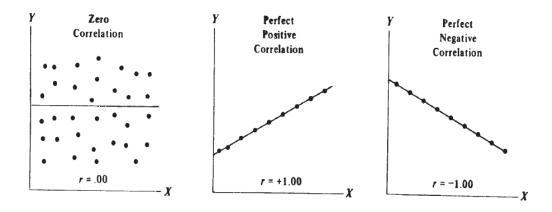


Figure 1: Scatter Plots which could be Linearly Regressed (SUSS 2016)

- Figure 1(A) shows a scatter plot that could be positively linearly regressed.
- Figure 1(B) shows a scatter plot that could be negatively linearly regressed.

The following scatter diagrams depict correlations of 0, +1.0, and -1.0.



# **3** | P A G E

#### II. STEP 1: SCATTER PLOT TO CHECK LINEARITY

Age (years) - X
1
1
2
2
3
3
4
4
5
5

Table 1: Repair Cost (\$) vs Age (years)

- Does the data in Table 1 show a linear relationship?
- We perform a scatter plot in Excel to find out.

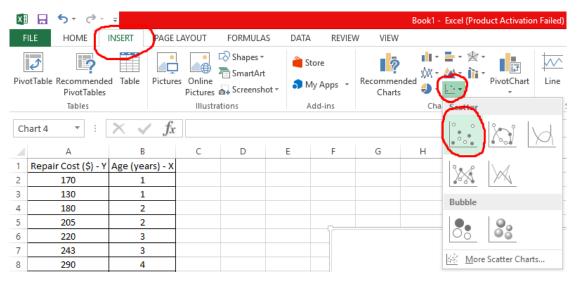


Figure 2: Scatter Plot

• Click Insert  $\rightarrow$  Choose the 1<sup>st</sup> Scatter Plot

# **4** | P A G E

XI	<u>□</u> 5· ∂·			5 1/0 1														?		5
<u>^</u>				<ul> <li>Excel (Produ</li> </ul>				_	ART TOOLS								-			
	HOME	INSERT PAGE	LAYOUT	FORMULA	S DAT	A REVIEW	/ VIEW	DESIGN	FORMA	Т					<u> </u>				ALVIN A	NG 👻 🔛
1.1		•											-	2	-	do	1			
		ange											-	th Row/h Sele						
	ent · Layout · Col													lumn Da						
С	nart Layouts					Cha	irt Styles							Data	Typ	e Locati	ion			
		× 4 f																		
D9	¥ :	$\times \checkmark f_{\lambda}$																		
	А	В	С	D	E	F	G	н	1	J	К	L	м	N	0	Р	Q	R	S	
1	Repair Cost (\$) - Y	Age (years) - X										_								
2	170	1	Select D	ata Source							? X									
3	130	1	Chart	data range:	1						<b>1</b>									
4	180	2												Empty Cha	art					
5	205	2																		
6	220	3			ſ	131		Column												
7	243	3		Entries (Serie:	d 🗸		н	rizontal (Cate	mon/ Avis Lal	hels										
8	290	4	ta /			K Remove		Edit	gory/rous car											
9	275	4			Equit /	V Wennove		E/ Luis												
10	404	5																		
11	380	5																		
12			-																	
13																				
14			_																	
15			Hidde	n and Empty	Cells					ОК	Cancel		_		_			_		
16 17			-																	

Figure 3: Label the Correct Axis

- Select the Empty Chart that popped up
- At the Design Tab, click "Select Data"
- Click "Add"

Repair Cost (\$) - Y	Age (years) - X	Edit Series ? ×
170	1	Savias nama:
130	1	Series <u>n</u> ame: =Sheet1!\$A\$1 <b>=</b> Repair Cost (\$
180	2	
205 🚤	2 🔫	Series X values:
220	3	= Sheet1!\$B\$2:\$B\$11 = 1, 1, 2, 2, 3,
243	3	Series <u>Y</u> values:
290	4	= Sheet1!\$A\$2:\$A\$11 = 170, 130, 180,
275	4	OK Cancel
404	5	
380	5	

- Series Name: Select "Repair Cost" header
- X Values: Select Age (years) column
- Y Values: Select Repair Cost (\$) column
- Click OK

# **5** | P A G E

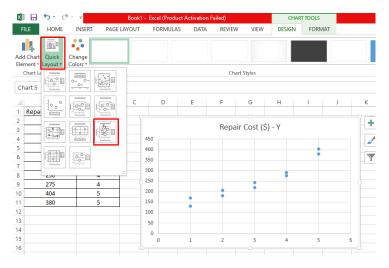


Figure 4: Editing using Quick Layout

• Click Quick Layout and select the fx graph.

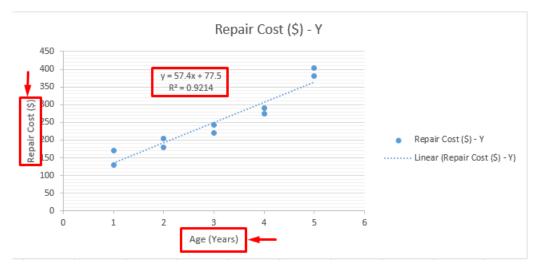


Figure 5: Final Scatter Plot with Linear Regression Line

- Figure 5 shows the final scatter plot with the linear regression line. This shows that the linearity check is OK.
- Remember to edit the Axis Titles.
- Note the Best Fit Line Equation given is  $\hat{Y} = 57.4X + 77.5$

# **6** | P A G E

#### ם לי פי <u>ה</u> י Book1 - Excel File Home Insert Page Layout Formulas Data Review View Add-ins Help $\ensuremath{\mathbb{Q}}$ Tell me what you want to do 💼 - 📕 - 🖧 -v 5 ٩ Ş 2 Get Add-ins <u>/// - ili - iìi</u> Ь PivotTable Recommended Table Illustrations Recommended Maps PivotChart Line Column Win/ My Add-ins 🧈 🗸 📴 🗸 -----PivotTables Charts Мар ~ Loss Tables Add-ins Tours Sparklines Repair Cost \$ Y ----м C Scatter Repair Cost \$ Y Age Years - X ۱. Use this chart type to: 170 1 J. · Compare at least two sets of 130 1 values or pairs of data. 0 0 180 Bu 2 Show relationships between sets 205 2 of values 220 3 6 243 3 Use it when: 290 4 The data represents separate 5 • measurements. 275 4 4 404 5 380 5 ¢₃ ¢ . . 2 14 15 1 16 17 0 0 50 100 150 200 250 300 350 400 450 ò -0 ò Sheet1 : 🔳 Sign in Page Layout Formulas Data View Add-ins Help Chart Design Format Q Tell me what you want to do File ık. Ş di. db Chart Lav Chart Styles Data Type Location *f*x Repair Cost \$ Y A1 Repair Cost \$ Y Age Years - X 170 1 Age Years - X Chart data range: =Sheet1!\$A\$1:\$B\$11 130 180 Switch Row/Column 205 5 220 Legend Entries (Seri tal (<u>C</u>ategory) Axis Label 243 4 •• 📩 Add 🛛 🔀 Edit X Bernove Edi<u>t</u> 290 275 Age Years - X 170 3 404 130 380 2 180 205 1 220

#### A. ALTERNATIVE WAY FOR CURVE FITTING

# 7 | PAGE

250 300

200

Hidden and Empty Cells

OK Cancel

0

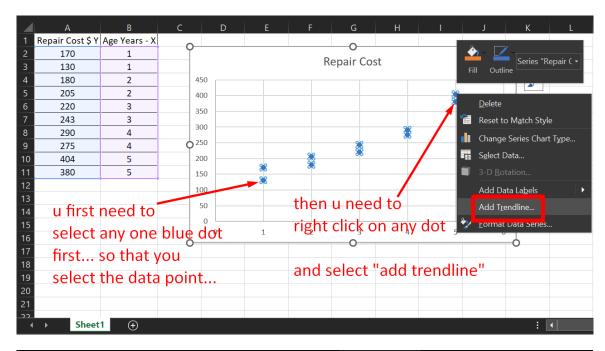
0 50 100 150

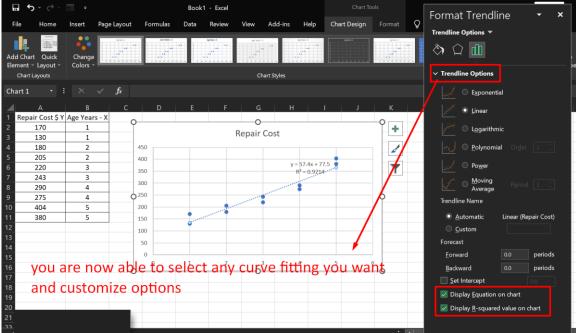
Select Data Source		×
Chart <u>d</u> ata range:		<u>↑</u>
Switch Row/Column		-
Legend Entries (Series) Horizontal (Category) Axis Labels		-
Add Free Add Transformed Add Transformed Add		
		-
		-
Hidden and Empty Cells OK	Car	ncel

A2	• :	×	$f_x$	Repa	ir Cost	\$ Y											
	А	В	с		D	E	F	G	н	I	J		к	L	М	N	О
1	Repair Cost \$ Y	Age Years - X															
2	170	1											Edit	Series			
3	130	1					H	epair Co	st				Series	name:			
4	180	2		450								7 [		- ir Cost		Ť	= Repair Cost
5	205	2		400							-			X values:			
6	220	3		350							•			et1!\$B\$2:\$B\$1	1	1	= 1, 1, 2, 2, 3,
7	243	3		300										Y values:			., ., _, _, _, _,
8	290	4							:					et1!\$A\$2:\$A\$1	1	1	= 170, 130, 180,
9	275	4		250									- one	CC1.9A92.9A91	1		- 170, 150, 100,
10	404	5		200												ОК	Cancel
1	380	5		150													
12				100													_
13				50													
14																	
15				0	0	1	2	3	4		5	6					
16					0	-	2	5	-			~					

	A	В	с	D	E	F	G	н		JJKLM	N O P O
1	Repair Cost \$ Y	Age Years - X								Select Data Source	
2	170	1					Repair C	`t			
3	130	1					Repair C	.ost		Chart data range:	1
4	180	2		450							ge is selected, it will replace all of the series in the Series panel.
5	205	2		400						Switch	Row/Column
6	220	3		350						V	~
7	243	3		300						Legend Entries (Series)	Horizontal (Category) Axis Labels
8	290	4							:	Add 🔀 Edit 🗙 Remove 🔺 🔻	Edit
9	275	4		250			:			Repair Cost	1
10	404	5		200		:	T				1
11	380	5		150						-	2
12				100							2
13				50							3
14										L	· · · · ·
15				0	1	2	3		4	Hidden and Empty Cells	OK Cancel
16				0	-	-	5				
17											

# **8** | P A G E





# **9** | P A G E

#### A - A X VIEW ALVIN ANG -DATA REVIEW 💾 Data Analysis 🔆 Clear **|**→□ **\***[] ns £↓ ZA Y **F** -• 87 ۳ę Reapply Filter Text to Flash Remove Data Consolidate What-If Relationships Group Ungroup Subtotal Z↓ Sort 🏷 Advanced Columns Fill Duplicates Validation Analysis Sort & Filter Data Tools alvsis ~ $\times$ Data Analysis ? Analysis Tools $\sim$ OK Covariance Descriptive Statistics Exponential Smoothing F-Test Two-Sample for Variances Fourier Analysis Histogram Moving Average Random Number Generation Pank and Percentile ^ Cancel F G н к S F Help Repair Cost (\$) - Y centile y = 57.4x + 77.5

**III.STEP 2: REGRESSION ANALYSIS** 

Figure 6: Click on Data Analysis

- In order to perform Regression Analysis, you first need to install Excel Analysis Toolpak.
- Please refer to Ang (2018) on how to install.
- Click the DATA tab  $\rightarrow$  Data Analysis  $\rightarrow$  Regression  $\rightarrow$  OK.

_		-	Let us	
1	Repair Cost (\$) - Y	Age (years) - X	Regression	? ×
2	170 🚽	1	Input	ОК
3	130	1	Input <u>Y</u> Range: SA\$1:SA\$11	
4	180	2	Input X Range: SB\$1:SB\$11	Cancel
5	205	2	3031.30311	Help
6	220	3	✓ Labels Constant is Zero	Helb
7	243	3	Confidence Level: 95 %	
8	290	4	Output options	
9	275	4		
10	404	5	O Triburnet	
11	380	5	New Worksheet Ply:	
12			O New Workbook	
13			Residuals	
14			Residuals Residual Plots	
15				
16			Normal Probability	
17			<u>N</u> ormal Probability Plots	
18				

Figure 7: Regression Analysis Inputs

- Input Y Range: Repair Cost column
- Input X Range: Age column
- Select Labels
- Select New Worksheet Ply
- Click OK

SUMMARY OUTPUT								
Regression St	atistics							
Multiple R	0.959911184							
R Square	0.92142948							
Adjusted R Square	0.911608165							
Standard Error	26.50212256							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	65895.2	65895.2	93.81936	1.07652E-05			
Residual	8	5618.9	702.3625					
Total	9	71514.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586	32.17664145	122.8233586
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678	43.73449322	71.06550678

Figure 8: Regression Analysis Output

- Figure 8 shows a new sheet created with all the Regression Analysis Output.
- We will explain it in detail in the next section.

#### IV. STEP 3: ANALYZING THE REGRESSION ANALYSIS OUTPUT

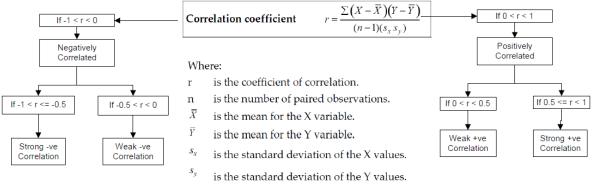
SUMMARY OUTPUT								
Regression St	atistics							
Multiple R	0.959911184							
R Square	0.92142948							
Adjusted R Square	0.911608165							
Standard Error	26.50212256							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	65895.2	65895.2	93.81936	1.07652E-05			
Residual	8	5618.9	702.3625					
Total	9	71514.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586	32.17664145	122.8233586
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678	43.73449322	71.06550678

#### A. MULTIPLE R

#### Figure 9: Multiple R

- Multiple R is also called : *Correlation Coefficient* & is also labelled as : r
- Multiple R or r shows the correlation between actual values of the dependent variable, Y and the predicted values for Y.
- Multiple  $\mathbf{R} = \mathbf{r} = 0$  indicates no correlation
- Multiple R = r = 1 means perfect correlation.
- Multiple R = r = 0.960 suggests a strong positive correlation.
- This means that as age of the car increases, so does annual repair cost.
- Figure 10 shows the equation of Multiple R or r.
- Since we rarely use calculation by hand, we shall ignore the equation. (Figure 11)

# **12** | P A G E



$$\overline{X} = \frac{\Sigma X}{n}$$
 and  $s = \sqrt{\frac{\Sigma (X - \overline{X})^2}{n-1}}$ 

	A	В	С	D	E	F	G
	Repair						
	Cost			Age			
	Ŷ	$(Y - \overline{Y})$	$(Y - \overline{Y})^2$	х	$(X-\overline{X})$	$(X-\overline{X})^2$	$(X-\overline{X})(Y-\overline{Y})$
	\$170	-80	6400	1	-2	4	160
	130	-120	14400	1	-2	4	240
	180	-70	4900	2	-1	1	70
	205	-45	2025	2	-1	1	45
	220	-30	900	3	0	0	0
	243	-7	49	3	0	0	0
	290	40	1600	4	1	1	40
	275	25	625	4	1	1	25
	404	154	23716	5	2	4	308
	380	130	16900	5	2	4	260
n	10			10			
Σ	2500	0.00	71515	30	0.00	20	1148

Figure 10: Multiple R = r = Correlation Coefficient

Step 1. Compute the means using sums in Column A and D:

$$\overline{Y} = \frac{\Sigma Y}{n} = \frac{2500}{10} = 250$$
  $\overline{X} = \frac{\Sigma X}{n} = \frac{30}{10} = 3.0$ 

Step 2. Compute the standard deviations using the sums in Column C and F:

$$s_y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n - 1}} = \sqrt{\frac{71515}{10 - 1}} = 89.14 \qquad \qquad s_x = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} = \sqrt{\frac{20}{10 - 1}} = 1.49$$

Step 3. Compute the coefficient of correlation *r* using the formula, the sum from Column G in the table, and the calculated standard deviations:

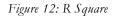
I	$r = \frac{\Sigma (X - \overline{X})(Y - \overline{Y})}{\Sigma - \overline{Y}}$	1148	= <u>1148</u> $=$ 0.9599 $=$ 0.960
I	$(n-1)(s_x s_y)$	9(1.4907)(89.14)	1196.71

Figure 11: How to Calculate Multiple R = r = by hand = not important

### **13** | P A G E

#### B. R SQUARE

SUMMARY OUTPUT								
Regression St	atistics							
Multiple R	0.959911184							
R Square	0.92142948							
Adjusted R Square	0.911608165							
Standard Error	26.50212256							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	65895.2	65895.2	93.81936	1.07652E-05			
Residual	8	5618.9	702.3625					
Total	9	71514.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586	32.17664145	122.8233586
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678	43.73449322	71.06550678



- R Square is known as the *Coefficient of Determination*
- In the case of Multiple Regression, it's called the Coefficient of *Multiple* Determination
- $R^2 = (Multiple R)^2 = r^2 = (0.9599)^2 = 0.9214$
- $R^2$  = the proportion of variation of Y accounted for by variation in X.
- For example, if  $R^2 = 0.92$ , that means that X (age) accounts for 92% of the variation of Y (repair cost).
- Since  $R^2 = \frac{SSR}{SST} = 1 \frac{SSE}{SST}$ , we shall describe what is SSR / SSE / SST in the next section.

С	SSR	1	SSE	1	SST
<b>U</b> .	001	/	001	/	001

SUMMARY OUTPUT										
Regression St	atistics									
Multiple R	0.959911	184								
R Square	0.92142	2948								
Adjusted R Square	0.911608	165								
Standard Error	26.50212	256								
Observations		10								
ANOVA										
	df		5	SS	MS	F	Significance F			
Regression		1	SSR	65895.2	SSR/165895.2	93.81936	1.07652E-05			
Residual	n-2	8	SSE	5618.9	702.3625	SSE/(n-2)				
Total	n-1	9	SST	71514.1						
	Coefficier	nts	Stando	ard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept		77.5	19.6	55450012	3.943117328	0.004277	32.17664145	122.8233586	32.17664145	122.8233586
age (years)	1	57.4	5.92	26054758	9.686039421	1.08E-05	43.73449322	71.06550678	43.73449322	71.06550678

Figure 13: SSR / SSE / SST ...

• SS: Sum of Squares

• SSR: Sum of Squares of Regression (variation) = 65895.2 = 
$$\sum (\hat{Y} - \overline{Y})^2$$

• Where  $\hat{Y} = 57.4X + 77.5$  (this equation was given by Excel Scatter Plot when we used the fx graph in Figure 5).

• Where 
$$\overline{Y} = \frac{\sum Y}{n} = 250$$
 (refer to Figure 11)

- We obtain  $\hat{Y} \& \overline{Y}$  through Table 2 below.
- SSE: Sum of Squares of Error (variation) = 5618.9 =  $\sum (Y \hat{Y})^2$
- SST: Sum of Squares Total (variation) = 71514.1 = = SSR + SSE =  $\sum (Y \overline{Y})^2$

**15** | P A G E

- MSR: Mean of Squares of Regression =  $702.3625 = \frac{SSR}{1}$
- MSE: Mean of Squares of Error =  $65895.2 = \frac{SSE}{n-2}$

	$\hat{Y} = 57.4$	X + 77.5			
X	bX	Ŷ	Y	$(Y-\hat{Y})$	$(Y-\hat{Y})^2$
1	57.40	134.9	\$170	35.1	1232.01
1	57.40	134.9	130	-4.9	24.01
2	114.80	192.3	180	-12.3	151.29
2	114.80	192.3	205	12.7	161.29
3	172.20	249.7	220	-29.7	882.09
3	172.20	249.7	243	-6.7	44.89
4	229.60	307.1	290	-17.1	292.41
4	229.60	307.1	278	-32.1	1030.41
5	287.00	364.5	404	39.5	1560.25
5	287.00	364.5	380	15.5	240.25
				SUM =	5618.9 <b>sse</b>

Table 2: Table to Obtain SSE / SSR / SST

#### D. DEGREES OF FREEDOM (DF)

SUMMARY OUTPUT							
Regression St	atistics						
Multiple R	0.959911	184					
R Square	0.92142	2948					
Adjusted R Square	0.911608	3165					
Standard Error	26.50212	2256					
Observations		10					
ANOVA							
	df		SS	MS	F	Significance F	
Regression		1	65895.2	65895.2	93.81936	1.07652E-05	
Residual	n-2	8	5618.9	702.3625			
Total	n-1	9	71514.1				
	Coefficie	nts	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept		77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586
age (years)		57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678

Figure 14: Degrees of Freedom (df)

- The Degree of Freedom (df) for Regression is 1
  - Reason:  $\hat{Y} = 57.4X + 77.5$
  - Y hat is dependent only on one X.
  - In simple words, there is only one way to get Y based on the Regression Line.
  - It means that I can predict every particular observation (Y: the regression line) based only on X
  - o That's why only 1 df.
- The df for Residual is 8
  - $\circ$  This means that there are 8/10 ways to get a guess of the residual.
  - The degrees of freedom associated with the error term (the residual) is (n 2)(SUSS 2016).

### **17** | P A G E

#### E. ADJUSTED R<sup>2</sup>

SUMMARY OUTPUT								
Regression St	atistics							
Multiple R	0.959911184							
R Square	0.92142948							
Adjusted R Square	0.911608165							
Standard Error	26.50212256							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	65895.2	65895.2	93.81936	1.07652E-05			
Residual	8	5618.9	702.3625					
Total	9	71514.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	77.5	19.65450012	3.943117328		32.17664145	122.8233586	32.17664145	122.823358
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678	43.73449322	71.0655067

Figure 15: Adjusted R Square

$$R_{adj}^{2} = 1 - \left[\frac{(1-R^{2})(n-1)}{n-k-1}\right]$$

Where:

- $R^{2}_{adj}$  is needed because  $R^{2}$  is not very accurate.
- $R^{2}_{adj}$  is more effective than R2.
- Because as the number of independent variables, X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub> etc... increases, R<sup>2</sup> increases.
- This makes R<sup>2</sup> inaccurate.
- $R^{2}_{adj}$  will not necessarily increase when a new variable is added.
- R<sup>2</sup><sub>adj</sub> takes into account the effect of potential overfitting due to the number of independent variables.
- R<sup>2</sup>: *Coefficient of Determination*

### **18** | P A G E

- n: number of observations = 10
- k: number of X's (or number of independent variables).
- In this case,  $\hat{Y} = 57.4X + 77.5$ , which means that there is only one X (age) (k = 1).
- In the case of Multiple Regression where there are many Xs, k will increase.

#### F. STANDARD ERROR

SUMMARY OUTPUT								
Degraceion Ct	atistics							
Regression St								
Multiple R	0.959911184							
R Square	0.92142948							
Adjusted R Square	0.911608165							
Standard Error	26.50212256							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	65895.2	65895.2	93.81936	1.07652E-05			
Residual	8	5618.9	702.3625					
Total	9	71514.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586	32.17664145	122.8233586
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678	43,73449322	71.06550678

Figure 16: Standard Error of Estimate

$$S_{y.x} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum \left(Y - \hat{Y}\right)^2}{n-2}}$$

- $S_{y.x}$ : Standard Error of Estimate = 26.5
- $S_{y,x}$  is a measure of dispersion of values around the regression line
- $S_{y,x}$ : is the Standard Deviation of the Residuals away from the proposed line.
- Figure 17 shows what Residuals are.

# **20** | P A G E

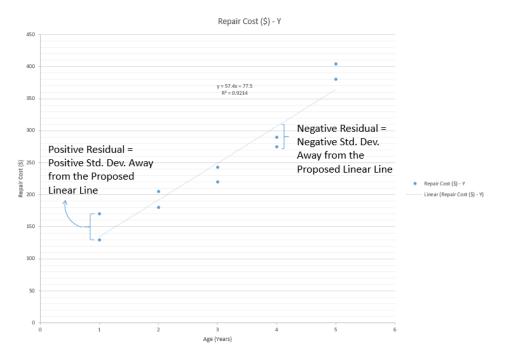
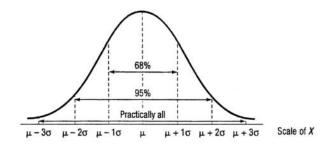


Figure 17: What are Residuals?

- Small  $S_{y,x}$  = Lesser Scatter = Good predictor
- Big  $S_{y,x}$  = More Scatter = Bad predictor
- Similar to Multiple R or r, both measures strength of relationship between X and Y
- But  $S_{y,x}$  has same units as Y, Multiple R or r has range -1 to 1
- Since S<sub>y.x</sub> = 26.5, this shows that about 68% of the predictions should be within ±\$26.50 (±1σ) of the actual repair costs and about 95% should be within (\$26.50 x 2) = ±\$53 (±2σ) of actual repair costs.



21 | P A G E

#### G. COEFFICIENTS

SUMMARY OUTPUT								
Regression St	atistics							
Multiple R	0.959911184							
R Square	0.92142948							
Adjusted R Square	0.911608165							
Standard Error	26.50212256							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	65895.2	65895.2	93.81936	1.07652E-05			
Residual	8	5618.9	702.3625					
Total	9	71514.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586	32.17664145	122.823358
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678	43.73449322	71.0655067

#### Figure 18: Coefficients

- Recall earlier that  $\hat{Y} = 57.4X + 77.5$
- The Y intercept (Repair Cost) = \$\$77.50.
- The gradient of X = 57.4.
- This shows that every increase of X (age) by 1 year
- $\rightarrow$  Y will increase linearly by Y = 57.4 (1) + 77.5 = \$134.90.

SUMMARY OUTPUT								
Regression St	atistics							
Multiple R	0.959911184							
R Square	0.92142948							
Adjusted R Square	0.911608165							
Standard Error	26.50212256							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	65895.2	65895.2	93.81936	1.07652E-05			
Residual	8	5618.9	702.3625					
Total	9	71514.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586	32.17664145	122.8233586
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678	43.73449322	71.06550678

Figure 19: F Stat vs t stat

• 
$$F = \frac{MSR}{MSE} = \frac{65895.2}{702.3625} = 93.819$$

- The purpose of F and Significance F is for Global Testing (to test for All X).
- In other words, the question (for multiple regression) is "Are ALL X important in this model?"
- Since this is a case of single variable linear regression, the question becomes "Is X (age) important in this model? Does Y (repair cost) really depend on it?"
- Let's create a Global Hypothesis test:
  - Null Hypothesis:  $H_0:\beta_1 = 0$  (X (age) is not important)
  - Alternate Hypothesis:  $H_1:\beta_1 \neq 0$  (X(age) is important)
  - Where  $\beta_1$  is actually referring to X.
  - $\circ \alpha = 5\%$

#### **23** | P A G E

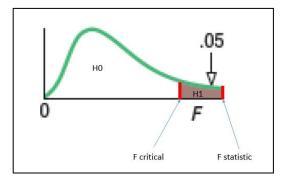
- Since df (Regression) = Numerator = 1
- Since df (Residual) = Denominator = 8
- Referring to Table 3, F critical  $(F_{crit}) = 5.32$
- Referring to Figure 19, F statistic  $(F_{stat}) = 93.82$
- Since  $F_{crit} < F_{stat} \rightarrow Accept H1 \rightarrow X$  is important.

Table 3: F distribution for alpha = 5%

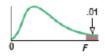
# **B.4 Critical Values of the** *F* **Distribution at a 5 Percent Level of Significance**

								)egrees o	f Freedo	m for the	e Numera	itor					
		1	2	3	4	5	6	7	8	0	10	12	15	20	24	30	40
	1 2 3 4 5	161 18.5 10.1 7.71 6.61	200 19.0 9.55 6.94 5.79	216 19.2 0.28 6.59 5.41	225 192 9.12 6.39 5.19	230 19,3 9,01 6,26 5,05	234 19.3 8.04 6.16 4.95	237 19.4 8.89 6.09 4.88	239 19.4 8.85 6.04 4.82	241 19.4 8.81 6.00 4.77	242 19.4 8.70 5.96 4.74	244 19.4 8.74 5.91 4.68	246 19,4 8,70 5,86 4,62	248 19.4 8.66 5.80 4.56	249 19.5 8.64 5.77 4.53	250 19.5 8.62 5.75 4.50	251 19.5 8.60 5.72 4.46
ter	6 7 9 10	5.99 5.59 5.32 5.12 4.96	5.14 4.74 4.46 4.26 4.10	4.76 4.35 4.07 3.86 3.71	4.53 4.12 3.84 3.63 3.48	4.39 3.07 3.69 3.48 3.33	4.28 3.87 3.58 3.37 3.22	4.21 3.79 3.50 3.29 3.14	4.15 3.73 3.44 3.23 3.07	4.10 3.69 3.39 3.18 3.02	4.06 3.64 3.35 3.14 2.98	4.00 3.57 3.28 3.07 2.91	3.94 3.51 3.22 3.01 2.85	3.87 3.44 3.15 2.94 2.77	3.84 3.41 3.12 2.90 2.74	3.81 3.38 3.08 2.86 2.70	3,77 3,34 3,04 2,83 2,66
fer the Denomina	11 12 13 14 15	4.84 4.75 4.67 4.60 4.54	3.98 3.89 3.81 3.74 3.68	3.59 3.49 3.41 3.34 3.29	3.36 3.26 3.18 3.11 3.06	3.20 3.11 3.03 2.96 2.90	3.09 3.00 2.92 2.95 2.79	3.01 2.91 2.83 2.76 2.71	2.95 2.85 2.77 2.70 2.64	2.90 2.90 2.71 2.65 2.59	2.85 2.75 2.67 2.60 2.54	2.79 2.69 2.60 2.53 2.48	2.72 2.62 2.53 2.46 2.40	2.65 2.54 2.46 2.39 2.33	2.61 2.51 2.42 2.35 2.29	2.57 2.47 2.38 2.31 2.25	2.53 2.43 2.34 2.27 2.20
Degrees of Freedom for the Denominator	16 17 18 19 20	4.49 4.45 4.41 4.38 4.35	3.63 3.59 3.55 3.52 3.49	3.24 3.20 3.16 3.13 3.10	3.01 2.96 2.93 2.90 2.90 2.87	2.85 2.81 2.77 2.74 2.71	2.74 2.70 2.66 2.63 2.60	2.66 2.61 2.58 2.54 2.51	2.59 2.55 2.61 2.48 2.45	2.54 2.49 2.46 2.42 2.39	249 245 241 238 235	2,42 2,38 2,34 2,31 2,28	2.35 2.31 2.27 2.23 2.20	2.28 2.23 2.19 2.16 2.12	2.24 2.19 2.15 2.11 2.08	210 215 211 207 204	215 210 206 203 1.00
Dec	21 22 23 24 25	4.32 4.30 4.28 4.26 4.24	3.47 3.44 3.42 3.40 3.39	3.07 3.05 3.03 3.01 2.99	2.84 2.82 2.90 2.78 2.76	2.68 2.68 2.64 2.62 2.60	2.57 2.55 2.53 2.51 2.49	2,49 2,46 2,44 2,42 2,40	2,42 2,40 2,37 2,36 2,34	2.37 2.34 2.32 2.30 2.29	232 230 227 225 224	2.25 2.23 2.20 2.18 2.16	2.18 2.15 2.13 2.11 2.09	2.10 2.07 2.05 2.03 2.01	2.05 2.03 2.01 1.98 1.96	2.01 1.98 1.96 1.94 1.92	1.96 1.94 1.91 1.89 1.87
	30 40 60 120 ∞	4.17 4.08 4.00 3.92 3.84	3.32 3.23 3.15 3.07 3.00	2.92 2.94 2.76 2.68 2.60	2.69 2.61 2.53 2.45 2.37	2.53 2.45 2.37 2.20 2.21	2.42 2.34 2.25 2.18 2.10	2.33 2.25 2.17 2.09 2.01	2.27 2.18 2.10 2.02 1.94	2.21 2.12 2.04 1.96 1.88	2.16 2.08 1.99 1.91 1.83	2.09 2.00 1.92 1.83 1.75	2.01 1.92 1.84 1.75 1.67	1.93 1.84 1.75 1.66 1.57	1.89 1.79 1.70 1.61 1.52	1.84 1.74 1.65 1.55 1.46	1.79 1.69 1.59 1.50 1.39

# **24** | P A G E



# **B.4** Critical Values of the *F* Distribution at a 1 Percent Level of Significance (*concluded*)



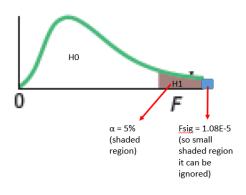
							legrees o	f Freedo	m for the	Numara	tor					
	1	2	3	4	5	6	7 7	8	0	10	12	15	20	24	30	40
-	4052	5000	5403	5625	5764	5959	5928	5991	6022	6056	6106	6157	6209	6235	6261	6297
	2 98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5
	3 34.1	30.B	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26,9	26.7	26.6	26.5	26.4
	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7
	5 16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.28	9.29
	13.7	10.9	9,78	9.15	8,75	8,47	B.26	8.10	7.98	7,87	7.72	7.56	7,40	7.31	7.23	7.14
1	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6,31	6.16	6.07	5.99	5,91
1	8 11.3	8.65	7.59	7.01	6.83	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12
1	9 10.6	8.02	6.99	6.42	6.06	5.90	5.61	5.47	5.35	5.26	5.11	4,96	4,81	4.73	4.65	4.57
j = 10	0 10.0	7.56	6.55	5.99	5.84	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17
Begrees of Freedom for the Denominator 2010년 11년 11년 11년 11년 11년 11년 11년 11년 11년	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86
5 13	2 9.33	6.93	5.95	5.41	5.08	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62
ē 1	3 9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43
<u>s</u> 14	8.86	6.51	5.56	5.04	4.89	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27
j 1	5 8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.99	3.90	3.67	3.52	3.37	3.29	3.21	3.13
5 10	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02
B 17	7 8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92
<u>분</u> 18		6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84
11	9 8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	276
8 20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.60
- 2		5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	264
22		5.72	4.B2	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58
23		5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	254
24		5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49
2	5 7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45
3		5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30
4(		5.18	4.31	3.83	3.51	3.29	3.12	2.99	2,89	2.80	2.66	2.52	2.37	2.29	2.20	211
6		4.98	4.13	3.85	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94
120		4.79	3.95	3.48	3.17	2.96	2.79	266	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76
	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.61	2.41	2.32	2.18	2.04	1.88	1.70	1.70	1.59

#### I. SIGNIFICANCE F

SUMMARY OUTPUT								
Regression St	atistics							
Multiple R	0.959911184							
R Square	0.92142948							
Adjusted R Square	0.911608165							
Standard Error	26.50212256							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	65895.2	65895.2	93.81936	1.07652E-05			
Residual	8	5618.9	702.3625					
Total	9	71514.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586	32.17664145	122.8233586
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678	43.73449322	71.06550678

#### Figure 20: Significance F

- Significance F  $(F_{sig})$  is just another method for doing F Stat.
- Similarly (following after F Stat), if  $\alpha = 5\%$
- & Significance F ( $F_{sig}$ ) = 1.08E-05
- $\rightarrow$  F<sub>sig</sub> <<  $\alpha$
- $\rightarrow$  Accept H1  $\rightarrow$  X is important.



# **26** | P A G E

J. T STAT

SUMMARY OUTPUT						
Regression St	atistics					
Multiple R	0.959911184					
R Square	0.92142948					
Adjusted R Square	0.911608165					
Standard Error	26.50212256					
Observations	10					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	65895.2	65895.2	93.81936	1.07652E-05	
Residual	8	5618.9	702.3625			
Total	9	71514.1				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678

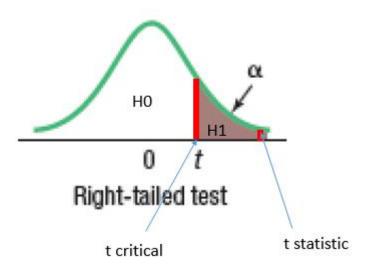
Figure 21: t stat

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.96\sqrt{10-2}}{\sqrt{1-(0.96)^2}} = \frac{2.7153}{0.28} = 9.690$$

- t stat = 9.686
- t critical = 1.86
  - o One-tailed test
  - $\circ \alpha = 0.05$  significance level
- Refer to Table 5: Student's t distribution below.
- Since;
  - Null Hypothesis:  $H_0:\beta_1 = 0$  (X (age) is not important)

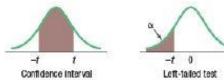
27 | PAGE

- Alternate Hypothesis:  $H_1: \beta_1 \neq 0$  (X(age) is important)
- t critical < t statistic  $\rightarrow$  H1 is accepted.

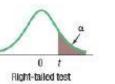


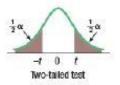
# **28** | PAGE

# **B.2 Student's t Distribution**









				e intervale						entidence in			
	\$0%	90%	95%	98%	99%	99.9%		80%	90%	95%	98%	99%	99.9%
		Lovel of	Significanc	e for One-T	ailed Test, o				Level of	Significanc	c fer Onc-T	ailed Test, a	
đ	0.10	0.05	0.025	0.01	0.005	0.0005	df	0.10	0.06	0.025	0.01	0.005	0.0008
		Lovel of	Significanc	e for Two-T	ailed Test, o			Ş	Level of	Significanc	e for Two-T	ailed Test, o	Ú.
	0.20	0.10	0.05	0.02	0.01	0.001		0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.708	31.821	63.657	638.619	36	1.306	1.688	2.028	2.434	2.719	3.582
2	1.886	2.920	4.303	6.965	9.925	31.599	37	1.305	1.687	2.028	2.431	2.715	3.574
3	1.838	2.353	3.182	4.541	5.841	12.924	38	1.304	1.686	2.024	2.429	2.712	3.566
4	1,533	2.132	2.776	3.747	4.604	8.610	39	1.304	1.685	2,023	2.428	2.708	3.558
5	1.476	2.015	2.571	3.365	4.032	6.869	40	1.303	1.684	2.021	2.423	2.704	3,551
6	1,440	1.943	2.447	3,143	3,707	5.959	41	1.303	1.683	2.020	2.421	2.701	3.544
7	1.415	1.895	2,365	2.998	3,499	5.408	42	1.302	1.682	2,018	2.418	2.698	3,539
8	1.397	1.860	2.306	2.896	3.355	5.041	43	1.302	1.681	2.017	2.416	2.695	3.532
9	1.383	1.833	2.262	2.821	3.250	4.781	44	1.301	1.680	2.015	2.414	2.692	3.526
10	1.372	1.812	2.228	2.764	3.160	4.587	45	1.301	1.679	2.014	2.412	2.690	3.520
11	1,363	1.796	2.201	2.718	3.106	4.437	46	1.300	1.679	2.013	2.410	2.687	3.515
12	1.356	1.782	2.179	2.681	3.055	4.318	47	1.300	1.678	2.012	2.408	2.685	3,510
13	1.350	1.771	2,160	2.650	3.012	4.221	48	1.290	1.677	2.011	2.407	2.682	3,505
14	1.345	1,761	2,145	2.624	2.977	4.140	49	1.299	1.677	2.010	2.405	2.680	3,500
15	1.341	1.753	2,131	2.602	2.947	4.073	50	1.299	1.676	2.009	2.403	2.678	3,496
16	1.337	1.746	2,120	2.583	2.921	4.015	51	1.298	1.675	2.008	2.402	2.676	3,492
17	1.333	1.740	2,110	2.567	2.898	3.965	52	1.298	1.675	2.007	2.400	2.674	3,489
18	1.330	1.734	2.101	2.552	2.878	3.922	53	1.298	1.674	2.006	2.399	2.672	3.484
19	1.328	1.729	2.093	2.539	2.861	3.883	54	1.297	1.674	2.005	2.397	2.670	3,480
20	1.325	1.725	2.086	2.528	2.845	3.850	55	1.297	1.673	2.004	2.396	2.668	3.476
21	1.323	1.721	2.080	2.518	2.831	3.819	56	1.297	1.673	2.003	2.395	2.667	3.473
22	1.321	1.717	2.074	2.508	2.810	3.792	57	1.297	1.672	2.002	2.394	2.665	3.470
23	1.310	1.714	2.060	2.500	2.807	3.768	58	1.296	1.672	2.002	2.392	2.663	3,488
24	1.318	1.711	2.064	2.492	2.797	3.745	59	1.296	1.671	2.001	2.391	2.682	3.483
25	1.316	1.706	2.060	2.485	2.787	3.725	60	1.296	1.671	2.000	2.390	2.660	3.480
26	1.315	1.706	2.056	2.470	2.779	3.707	61	1.296	1.670	2.000	2.380	2.659	3.457
27	1.314	1.703	2.052	2.473	2.771	3.690	62	1.295	1.670	1.900	2.388	2.657	3.454
28	1.313	1.701	2.048	2.467	2.763	3.674	63	1.295	1.669	1.998	2.387	2.656	3,452
29	1.311	1.609	2.045	2.462	2.756	3.659	64	1.295	1.669	1.998	2.386	2.655	3.449
30	1.310	1.607	2.042	2,457	2.750	3.646	65	1.295	1.660	1.907	2.385	2.654	3.447
31	1.309	1.696	2.040	2.453	2.744	3.633	66	1.295	1.668	1.997	2.384	2.652	3.444
32	1.300	1.604	2.037	2,440	2.738	3.622	67	1.294	1.668	1.906	2.383	2.651	3.442
33	1.308	1.802	2.035	2.445	2.733	3.611	68	1.294	1.668	1.905	2.382	2.650	3.430
34	1.307	1.601	2.032	2.441	2.728	3.601	69	1.294	1.667	1.995	2.382	2.649	3.437
35	1.306	1.890	2.030	2.438	2.724	3.591	70	1.294	1.667	1.904	2.381	2.648	3.435

(continued)

# **29** | P A G E

Confidence Intervals, c								Co	mfidence in	tervals, c			
	80%	90%	95%	96%	99%	99.9%	1	80%	90%	95%	98%	99%	99.9%
Level of Significance for One-Tailed Test, $\alpha$							Level of Significance for One-Tailed Test, $\alpha$						
df	0.10	0.05	0.025	0.01	0.005	0.0005	df	0.10	0.05	0.025	0.01	0.005	0.0005
	8 S	Level of	Significance	e for Two-Ta	nileci Test, a	2		E - 8	Level of	Significanc	e for Two-T	alled Test, o	2
	0.20	0.10	0.05	0.02	0.01	0.001		0.20	0.10	0.05	0.02	0.01	0.001
71	1.294	1.667	1.994	2.380	2.647	3.433	89	1.291	1.662	1.987	2.369	2.632	3,403
72	1.293	1.656	1.993	2.379	2.848	3.431	90	1.291	1.862	1.987	2.368	2.632	3,402
73	1.293	1.666	1.993	2.379	2.645	3.429	842201	100000	1001000000	100000000	0100-001	11992200	0.000
74	1.293	1.666	1.993	2.378	2.644	3.427	91	1.291	1.662	1.986	2.368	2.631	3,401
75	1.293	1.665	1.992	2.377	2.643	3.425	92	1.291	1.662	1.986	2.368	2,630	3.399
	100 410 A.S. S		2012/00/00		6.57 Deem		93	1.291	1.661	1.986	2.367	2.630	3,396
76	1.293	1,665	1.992	2.376	2.642	3.423	94	1.291	1.661	1.986	2.367	2,629	3.397
77	1.293	1.665	1.991	2.376	2.641	3.421	96	1.291	1.661	1.985	2.366	2.629	3.396
78	1.292	1.665	1.991	2.375	2.640	3.420		3387			1222	10000	
79	1.292	1.664	1.990	2.374	2.640	3.418	96	1.290	1.661	1.985	2.386	2.628	3.395
80	1.292	1.664	1.990	2.374	2.639	3.416	97	1,290	1.661	1.985	2.365	2.627	3.394
							98	1.290	1.661	1.984	2.365	2.627	3.393
81	1.292	1.664	1.990	2.373	2.838	3.415	99	1.290	1.660	1.984	2.365	2.626	3.392
82	1.292	1.664	1.989	2.373	2.637	3.413	100	1.290	1.660	1.984	2.364	2,626	3.390
83	1.292	1.663	1.989	2.372	2.636	3.412		12233	23221	1993		12232	1223
84	1.292	1.663	1.989	2.372	2.636	3.410	120	1.289	1.658	1,980	2.358	2.617	3.373
85	1.292	1.663	1.988	2.371	2.635	3.409	140	1.288	1.658	1.977	2.553	2.611	3.361
133	12.238	0330	3.5439	333	100000	133233	160	1.287	1,654	1.975	2.350	2,607	3,352
86	1.291	1.663	1.988	2.370	2.634	3.407	180	1.286	1.653	1.973	2.347	2.603	3,345
87	1.291	1.663	1.988	2.370	2.634	3.406	200	1.286	1.653	1.972	2.345	2,601	3.340
88	1.291	1.662	1.997	2.369	2.633	3.405	90	1.282	1,645	1.960	2.326	2,576	3,291

#### K. P-VALUE

age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678	43.73449322	71.06550678
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586	32.17664145	122.8233586
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Total	9	71514.1						
Residual	8	5618.9	702.3625					
Regression	1	65895.2	65895.2	93.81936	1.07652E-05			
	df	SS	MS	F	Significance F			
ANOVA								
Observations	10							
Standard Error	26.50212256							
Adjusted R Square	0.911608165							
R Square	0.92142948							
Multiple R	0.959911184							
Regression St	atistics							
SUMMARY OUTPUT								

Figure 22: P Value

- Notice P-Value for X (Age) = 1.08E-05 is exactly the same as Significance F ( $F_{sig}$ ) = 1.08E-05• 05
- This is because there is only one X (age). •
- If there are multiple X's, they will be different. •
- Since P-Value for X (Age) =  $1.08\text{E}-05 \ll \alpha = 5\%$ •
- Likewise accept H1  $\rightarrow$  X is important. •

#### L. LOWER AND UPPER 95%

SUMMARY OUTPUT						
Regression St	atistics					
Multiple R	0.959911184					
R Square	0.92142948					
Adjusted R Square	0.911608165					
Standard Error	26.50212256					
Observations	10					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	65895.2	65895.2	93.81936	1.07652E-05	
Residual	8	5618.9	702.3625			
Total	9	71514.1				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	77.5	19.65450012	3.943117328	0.004277	32.17664145	122.8233586
age (years)	57.4	5.926054758	9.686039421	1.08E-05	43.73449322	71.06550678

Figure 23: Lower and Upper 95%

- How did we get Lower 95% = 43.73 and Upper 95% = 71.066 for Age?
- The Mean Coefficient is 57.4.
- The 95% CI is = 57.4 43.73 = 13.67.
- Thus, the 95% CI is 57.4  $\pm$  13.67  $\rightarrow$  (43.73, 71.07).

- 1. X values are independent.
- 2. Y is dependent on X.
- 3. Y values are Normally Distributed
- 4. Means of Y values lie on the Regression Line.
- 5.  $S_{y,x}$  is the Std. Dev. of these Y values

 $\hat{Y} \pm 1 s_{y,x}$  encompasses about 68% of the observed values.

 $\hat{Y} \pm 2s_{y,x}$  encompasses about 95% of the observed values.

 $\hat{Y} \pm 3s_{y,x}$  encompasses virtually all of the observed values.

- 6.  $S_{y,x}$  is a fixed constant.
- 7. There's no relationship between each Y value.
- 8. Each X has been picked independent of another X.

#### VI. LINEAR REGRESSION BY HAND

• Linear Regression by hand is also known as the "Method of Least Squares".

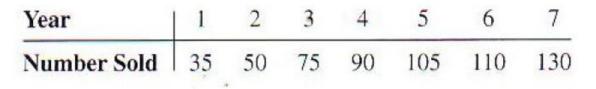


Figure 24: Example for Method of Least Squares

• Presume we want to obtain the Linear Equation for Figure 24.

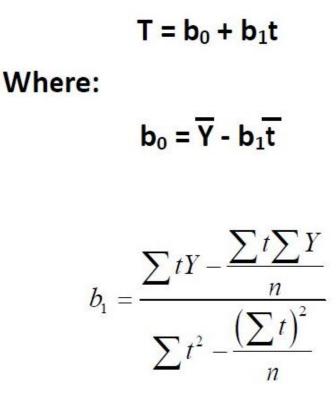
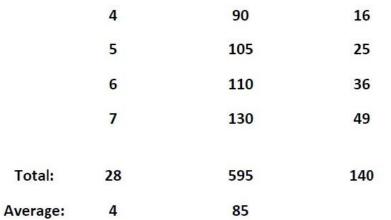


Figure 25: Formula for Method of Least Squares

### **34** | P A G E

	Year (t)	Numbers Sold (Y)	t2	tY
	1	35	1	35
	2	50	4	100
	3	75	9	225
	4	90	16	360
	5	105	25	525
	6	110	36	660
	7	130	49	910
ıl:	28	595	140	2815
ge:	4	85		

Table 6: Applying the Method of Least Squares



$$b_{1} = \frac{\sum tY - \frac{\sum t\sum Y}{n}}{\sum t^{2} - \frac{(\sum t)^{2}}{n}}$$

$$= \frac{2815 - \frac{(28)(595)}{7}}{140 - \frac{28^{2}}{7}}$$

$$= 15.5357$$

$$b_{0} = Y - b_{1}t$$

$$= 85 - 15.5357 (4)$$

$$= 22.857$$
Hence Trend Line Equation:  

$$T = b_{0} + b_{1}t$$

$$T = 22.857 + 15.536t$$

**35** | P A G E

#### ABOUT DR. ALVIN ANG



Dr. Alvin Ang earned his Ph.D., Masters and Bachelor degrees from NTU, Singapore. He used to be a Scientist and Professor. He is currently an Entrepreneur and Business Consultant. More about him at his website <u>www.AlvinAng.sg</u>.

# **36** | P A G E